A SCALE INVARIANT BAYESIAN METHOD TO SOLVE LINEAR INVERSE PROBLEMS

Ali Mohammad-Djafari and Jérôme Idier
Laboratoire des Signaux et Systèmes (CNRS-ESE-UPS)
École Supérieure d'Électricité,
Plateau de Moulon, 91192 Gif-sur-Yvette Cédex, France

Abstract. In this paper we propose a new Bayesian estimation method to solve linear inverse problems in signal and image restoration and reconstruction problems which has the property to be scale invariant. In general, Bayesian estimators are *nonlinear* functions of the observed data. The only exception is the Gaussian case. When dealing with linear inverse problems the linearity is sometimes a too strong property, while *scale invariance* often remains a desirable property. As everybody knows one of the main difficulties with using the Bayesian approach in real applications is the assignment of the direct (prior) probability laws before applying the Bayes' rule. We discuss here how to choose prior laws to obtain scale invariant Bayesian estimators. In this paper we discuss and propose a familly of generalized exponential probability distributions functions for the direct probabilities (the prior $p(\boldsymbol{x})$ and the likelihood $p(\boldsymbol{y}|\boldsymbol{x})$), for which the posterior $p(\boldsymbol{x}|\boldsymbol{y})$, and, consequently, the main posterior estimators are scale invariant. Among many properties, generalized exponential can be considered as the maximum entropy probability distributions subject to the knowledge of a finite set of expectation values of some knwon functions.

1. Introduction

We address a class of linear inverse problems arising in signal and image reconstruction and restoration problems which is to solve integral equations of the form:

$$g_{ij} = \int_D f(\boldsymbol{r}')\, h_{ij}(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}' + b_{ij}, \quad i,j = 1,\cdots,M, \tag{1}$$

where $\boldsymbol{r}' \in \mathbb{R}^2$, $f(\boldsymbol{r}')$ is the object (image reconstruction problems) or the original image (image restoration problems), g_{ij} are the measured data (the projections in image reconstruction or the degraded image in image restoration problems), b_{ij} are the measurement noise samples and $h_{ij}(\boldsymbol{r}')$ are known functions which depend only on the measurement system. To show the generality of this relation, we give in the following some applications we are interested in:

– Image restoration:

$$g(x_i,y_j) = \iint_D f(x',y')h(x_i - x', y_j - y')\,\mathrm{d}x'\,\mathrm{d}y' + b(x_i,y_j) \quad , \quad \begin{array}{l} i = 1,\cdots,N \\ j = 1,\cdots,M \end{array} ,$$

where $g(x_i,y_j)$ are the observed degraded image pixels and $h(x,y)$ is the point spread function (PSF) of the measurement system.

– X-ray computed tomography (CT):

$$g(r_i,\phi_j)=\iint_D f(x,y)\delta(r_i-x\cos\phi_i-y\sin\phi_i)\,\mathrm{d}x\,\mathrm{d}y+b(r_i,\phi_j) \quad , \quad \begin{array}{l} i=1,\cdots,N \\ j=1,\cdots,M \end{array} \; ,$$

where $g(r_i,\phi_j)$ are the projections along the axis $r_i = x\cos\phi_i - y\sin\phi_i$, having the angle ϕ_j, and which can be considered as the samples of the Radon transform (RT) of the object function $f(x,y)$.

– Fourier Synthesis in radio astronomy, in SAR imaging and in diffracted wave tomographic imaging systems:

$$g(u_j,v_j)=\iint_D f(x,y)\,\exp\left[-j(u_jx+v_jy)\right]\,\mathrm{d}x\,\mathrm{d}y+b(u_j,v_j), \quad j=1,\cdots,M,$$

where $\boldsymbol{u}_j=(u_j,v_j)$ is a radial direction and $g(u_j,v_j)$ are the samples of the complex valued visibility function of the sky in radio astronomy or the Fourier transform of the measured signal in SAR imaging.

Other examples can be found in [6, 7, 5, 8, 9].

In all these applications we have to solve the following ill-posed problem: how to estimate the function $f(x,y)$ from some finite set of measured data which may also be noisy, because there is no experimental measurement device, even the most elaborate, which could be entirely free from uncertainty, the simplest example being the finite precision of the measurements.

The numerical solution of these equations needs a discretization procedure which can be done by a quadrature method. The linear system of equations resulting from the discretization of an ill-posed problem is, in general, very ill-conditioned if not singular. So the problem is to find a unique and stable solution for this linear system. The general methods which permit us to find a unique and stable solution to an ill-posed problem by introducing an *a priori* information on the solution are called regularization . The *a priori* information can be either in a deterministic form (positivity) or in a stochastic form (some constraints on the probability density functions).

When discretized, these problems can be described by the following:

> "Estimate a vector of the parameters $\boldsymbol{x}\in\mathbb{R}^n$ (pixel intensities in an image for example) given a vector of measurements $\boldsymbol{y}\in\mathbb{R}^m$ (representing, for example, either a degraded image pixel values in restoration problems or the projections values in reconstruction problems) and a linear transformation $\boldsymbol{A}$ relating them by:
>
> $$\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}, \tag{2}$$
>
> where $\boldsymbol{b}$ represents the discretization errors and the measurement noise which is supposed to be zero-mean and additive."

In this paper we propose to use the Bayesian approach to find a regularized solution to this problem. Noting that the Bayesian theory only gives us a framework for the formulation of the inverse problem, not a solution of it. The main difficulty is, in general, before the application of the Bayes' formula, *i.e.*; how to formulate appropriately the problem and how to assign the direct probabilities. Keeping this fact in mind, we propose the following organization to this paper: In section 2. we give a brief description of the Bayesian approach with detail calculations of the solution in the special case of Gaussian laws. In section 3. we discuss about the *scale invariance* property and propose a familly of prior probability density functions (*pdf*) which insure this property for the solution. Finally, in section 4., we present some special cases and give detailed calculations for the solution.

2. General Bayesian approach

A general Bayesian approach involves the following steps:

- Assign a prior probability law $p(\boldsymbol{x})$ to the unknown parameter to translate our incomplete *a priori* information (prior beliefs) about these parameters;
- Assign a direct probability law to the measured data $p(\boldsymbol{y}|\boldsymbol{x})$ to translate the lack of total precision and the inevitable existence of the measurement noise;
- Use the Bayes' rule to calculate the posterior law $p(\boldsymbol{x}|\boldsymbol{y})$ of the unknown parameters;
- Define a decision rule to give values $\widehat{\boldsymbol{x}}$ to these parameters.

To illustrate the whole procedure, let us to consider an example; the Gaussian case. If we suppose that what we know about the unknown input $\boldsymbol{x}$ is its mean $\mathrm{E}\left\{\boldsymbol{x}\right\} = \boldsymbol{x}_0$ and its covariance matrix $\mathrm{E}\left\{(\boldsymbol{x}-\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0)^t\right\} = \boldsymbol{R}_x = \sigma_x^2 \boldsymbol{P}$, and what we know about the measurement noise $\boldsymbol{b}$ is also its covariance matrix $\mathrm{E}\left\{\boldsymbol{b}\boldsymbol{b}^t\right\} = \boldsymbol{R}_b = \sigma_b^2 \boldsymbol{I}$, then we can use the maximum entropy principle to assign:

$$p(\boldsymbol{x}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^t {\boldsymbol{R}_x}^{-1}(\boldsymbol{x}-\boldsymbol{x}_0)\right], \tag{3}$$

and

$$p(\boldsymbol{y}|\boldsymbol{x}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^t {\boldsymbol{R}_b}^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})\right]. \tag{4}$$

Now we can use the Bayes' rule to find:

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{x}), \tag{5}$$

and use, for example, the maximum a posteriori (MAP) estimation rule to give a solution to the problem, *i.e.*;

$$\widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \left\{p(\boldsymbol{x}|\boldsymbol{y})\right\}, \tag{6}$$

Other estimators are possible. In fact, all we want to know is resumed in the posterior law. In general, one can construct a bayesian estimator by defining a cost (or utility) function $C(\widehat{\boldsymbol{x}}, \boldsymbol{x})$ and by minimizing its mean value

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{z}} \left\{\mathrm{E}_{X|Y}\left\{C(\boldsymbol{z},\boldsymbol{x})\right\}\right\} = \arg\min_{\boldsymbol{z}} \left\{\int C(\boldsymbol{z},\boldsymbol{x}) p(\boldsymbol{x}|\boldsymbol{y})\, \mathrm{d}\boldsymbol{x}\right\}.$$

The two classical estimators:

– Posterior mean (PM): $\quad \widehat{\boldsymbol{x}} = \mathrm{E}_{X|Y}\{\boldsymbol{x}\} = \displaystyle\int \boldsymbol{x}\, p(\boldsymbol{x}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}$,
  is obtained when defining $C(\widehat{\boldsymbol{x}}, \boldsymbol{x}) = (\widehat{\boldsymbol{x}} - \boldsymbol{x})^t(\widehat{\boldsymbol{x}} - \boldsymbol{x})$, and
– Maximum *a posteriori* (MAP): $\quad \widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{y})\}$,
  is obtained when defining $C(\widehat{\boldsymbol{x}}, \boldsymbol{x}) = 1 - \delta(\widehat{\boldsymbol{x}} - \boldsymbol{x})$.

Now, let us go a little further inside the calculations. Replacing (3), and (4) in (5), we calculate the posterior law:

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left[-\frac{1}{2\sigma_b^2} J(\boldsymbol{x})\right], \quad \text{with } J(\boldsymbol{x}) = (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^t(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) + \lambda(\boldsymbol{x} - \boldsymbol{x}_0)^t \boldsymbol{P}^{-1}(\boldsymbol{y} - \boldsymbol{x}_0),$$

where $\lambda = \sigma_b^2/\sigma_x^2$. The posterior is then also a Gaussian. We can know use any decision rule to obtain a solution. For example the maximum a posteriori (MAP) solution is obtained by:

$$\widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{y})\} = \arg\min_{\boldsymbol{x}} \{J(\boldsymbol{x})\}\,. \tag{7}$$

Note that in this special Gaussian case both estimators, *i.e.*; the posterior mean (PM) and the MAP estimators are the same:

$$\widehat{\boldsymbol{x}} = \mathrm{E}_{X|Y}\{\boldsymbol{x}\} = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{y})\} \tag{8}$$

and the minimization of the criterion $J(\boldsymbol{x})$ which can also be written in the form:

$$J(\boldsymbol{x}) = ||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||^2 + \lambda||\boldsymbol{x} - \boldsymbol{x}_0||^2_{\boldsymbol{P}} \tag{9}$$

can be considered as a regularization procedure to the inverse problem (2). Indeed, the Bayesian approach will give us here a new interpretation of the regularization parameter in terms of the signal to noise ratio, *i.e.*; $\lambda = \sigma_b^2/\sigma_x^2$.

$J(\boldsymbol{x})$ is a quadratic function of $\boldsymbol{x}$. The solution $\widehat{\boldsymbol{x}}$ is then a linear function of the data $\boldsymbol{y}$. This is due to the fact that the problem is linear and all the probability laws are Gaussian. Excepted this case, in general, the Bayesian estimators are not linear functions of the observations $\boldsymbol{y}$. However, we may not need that the solution be a linear function of the data $\boldsymbol{y}$, but the *scale invariance* is the minimum property which is often needed.

3. Scale invariant Bayesian estimators

What we are proposing in this paper is to study in what conditions we can obtain estimators who are scale invariant. Note that *linearity* is the combination of

$$\textit{additivity:} \qquad \left\{\begin{array}{l} \boldsymbol{y}_1 \mapsto \widehat{\boldsymbol{x}}_1, \\ \boldsymbol{y}_2 \mapsto \widehat{\boldsymbol{x}}_2 \end{array}\right. \Longrightarrow \boldsymbol{y}_1 + \boldsymbol{y}_2 \mapsto \widehat{\boldsymbol{x}}_1 + \widehat{\boldsymbol{x}}_2,$$

and

$$\textit{scale invariance:} \qquad \boldsymbol{y}_1 \mapsto \widehat{\boldsymbol{x}}_1 \Longrightarrow \forall k > 0,\ k\boldsymbol{y}_1 \mapsto k\widehat{\boldsymbol{x}}_1.$$

In a linear inverse problem what is often necessary is that the solution be scale invariant. As we have seen in the last section when all the probability laws are Gaussian then the Bayesian estimators are linear functions of the data, so that the methods based on this assumption have not to take care about the scale of the measured data. The Gaussian assumption is very restrictive. On the other hand, more general priors yield the Bayesian estimators which are nonlinear functions of data, so the result of the inversion method depend on the absolute values of the measured data. In other words, two users of the method using two different scale factors would not get the same results, even rescaled:

$$\begin{array}{lc} \boldsymbol{y} \longrightarrow \boxed{k_1} \longrightarrow \boxed{\text{Estimation}} \longrightarrow \widehat{\boldsymbol{x}}_1 & \\ & \frac{\widehat{\boldsymbol{x}}_2}{k_2} \neq \frac{\widehat{\boldsymbol{x}}_1}{k_1} \\ \boldsymbol{y} \longrightarrow \boxed{k_2} \longrightarrow \boxed{\text{Estimation}} \longrightarrow \widehat{\boldsymbol{x}}_2 & \end{array}$$

A general nonlinear (scale variant) estimation method

What we want to specify in this paper is a family of probability laws for which these estimators are scale invariant. So the user of the inversion method can process the data without worrying about rescaling them to an arbitrary level and two users of the method at two different scales will obtain the proportional results:

$$\begin{array}{lc} \boldsymbol{y} \longrightarrow \boxed{k_1} \longrightarrow \boxed{\text{Estimation}} \longrightarrow \widehat{\boldsymbol{x}}_1 & \\ & \frac{\widehat{\boldsymbol{x}}_2}{k_2} = \frac{\widehat{\boldsymbol{x}}_1}{k_1} \\ \boldsymbol{y} \longrightarrow \boxed{k_2} \longrightarrow \boxed{\text{Estimation}} \longrightarrow \widehat{\boldsymbol{x}}_2 & \end{array}$$

A scale invariant estimation method

To do this let us note

- $\boldsymbol{\theta}$ all the unknown parameters defining our measuring system (noise variance σ^2 and the prior law parameters for example),
- $p_1(\boldsymbol{x}_1|\boldsymbol{y}_1;\boldsymbol{\theta}_1)$ and $p_k(\boldsymbol{x}_k|\boldsymbol{y}_k;\boldsymbol{\theta}_k)$ the two expressions of the posterior law for scale 1 and for scale k with

$$\boldsymbol{x}_k = k\boldsymbol{x}_1, \quad \boldsymbol{y}_k = k\boldsymbol{y}_1.$$

Then, what we need is the following:

$$\exists \boldsymbol{\theta}_k = f(\boldsymbol{\theta}_1, k) \,|\, \forall k > 0, \forall \boldsymbol{x}_1, \boldsymbol{y}_1, \quad p_k\left(\boldsymbol{x}_k|\boldsymbol{y}_k;\boldsymbol{\theta}_k\right) = \frac{1}{k^n}\, p_1(\boldsymbol{x}_1|\boldsymbol{y}_1;\boldsymbol{\theta}_1), \tag{10}$$

which means that the functional form of the posterior law remains unchanged when the measurement's scale is changed. Only we have to modify the parameters $\boldsymbol{\theta}_k = f(\boldsymbol{\theta}_1, k)$ which is only a function of $\boldsymbol{\theta}_1$ and the scale factor k.

However, not all estimators based on this posterior will be scale invariant. The cost function must also have some property to obtain a scale invariant estimator. So, the main result of this paper can be resumed in the following theorem:

Theorem: If $\exists \boldsymbol{\theta}_k = f(\boldsymbol{\theta}_1, k) \,|\, \forall k > 0, \forall \boldsymbol{x}_1, \boldsymbol{y}_1$,

$$p_k\left(\boldsymbol{x}_k|\boldsymbol{y}_k;\boldsymbol{\theta}_k\right) = \frac{1}{k^n}\, p_1(\boldsymbol{x}_1|\boldsymbol{y}_1;\boldsymbol{\theta}_1),$$

then any bayesian estimator with a cost function $C(\widehat{\boldsymbol{x}}, \boldsymbol{x})$ satisfying:

$$C(\widehat{\boldsymbol{x}}_k, \boldsymbol{x}_k) = a_k + b_k C(\widehat{\boldsymbol{x}}, \boldsymbol{x}),$$

is a scale invariant estimator, *i.e.*;

$$\widehat{\boldsymbol{x}}_k(\boldsymbol{y}_k;\boldsymbol{\theta}_k) = k\widehat{\boldsymbol{x}}_1(\boldsymbol{y}_1;\boldsymbol{\theta}_1).$$

Proof: In fact, it is easy to see the following:

$$\begin{aligned}
\widehat{\boldsymbol{x}}_k(\boldsymbol{y}_k;\boldsymbol{\theta}_k) \quad &= \arg\min_{\boldsymbol{z}_k}\left\{\int C(\boldsymbol{z}_k,\boldsymbol{x}_k) p_k(\boldsymbol{x}_k|\boldsymbol{y}_k;\boldsymbol{\theta}_k)\,\mathrm{d}\boldsymbol{x}_k\right\}\\
&= k\arg\min_{\boldsymbol{z}_1}\left\{\int [b_k C(\boldsymbol{z}_1,\boldsymbol{x}_1)+a_k]\frac{1}{k^n}\, p_1(\boldsymbol{x}_1|\boldsymbol{y}_1;\boldsymbol{\theta}_1) k^n\,\mathrm{d}\boldsymbol{x}_1\right\}\\
&= k\arg\min_{\boldsymbol{z}_1}\left\{b_k \int C(\boldsymbol{z}_1,\boldsymbol{x}_1) p_1(\boldsymbol{x}_1|\boldsymbol{y}_1;\boldsymbol{\theta}_1)\,\mathrm{d}\boldsymbol{x}_1 + a_k\right\}\\
&= k\arg\min_{\boldsymbol{z}_1}\left\{\int C(\boldsymbol{z}_1,\boldsymbol{x}_1) p_1(\boldsymbol{x}_1|\boldsymbol{y}_1;\boldsymbol{\theta}_1)\,\mathrm{d}\boldsymbol{x}_1\right\}\\
&= k\,\widehat{\boldsymbol{x}}_1(\boldsymbol{y}_1;\boldsymbol{\theta}_1)
\end{aligned}$$

Note the great significance of this result, even if the estimateur $\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\theta})$ is a nonlinear function of the observations $\boldsymbol{y}$ it stays scale invariant.

Now, the task is to search for a large familly of probability laws $p(\boldsymbol{x})$ and $p(\boldsymbol{y}|\boldsymbol{x})$ in a manner that the posterior law $p(\boldsymbol{x}|\boldsymbol{y})$ remains scale invariant. We propose to do this search in the generalized exponential familly for two reasons:

- First the generalized exponential probability density functions form a very rich one, and
- Second, they can be considered as the maximum entropy prior laws subject to a finite number of constraints (linear or nonlinear).

Noting also that if $p(\boldsymbol{x})$ and $p(\boldsymbol{y}|\boldsymbol{x})$ are scale invariant then the posterior $p(\boldsymbol{x}|\boldsymbol{y})$ is also scale invariant and that there is a symmetry for $p(\boldsymbol{x})$ and $p(\boldsymbol{y}|\boldsymbol{x})$, so that it is only necessary to find the scale invariance conditions for one of them. In the following, without loss of generality, we consider the case where $p(\boldsymbol{y}|\boldsymbol{x})$ is Gaussian:

$$p(\boldsymbol{y}|\boldsymbol{x};\sigma^2) \propto \exp\left[-\chi^2(\boldsymbol{x},\boldsymbol{y};\sigma^2)\right], \quad \text{with } \chi^2(\boldsymbol{x},\boldsymbol{y};\sigma^2) = \frac{1}{2\sigma^2}\,[\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}]^t[\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}], \tag{11}$$

and find the conditions for $p(\boldsymbol{x})$ to be scale invariant. We choose the generalized exponential *pdf*'s for $p(\boldsymbol{x})$, *i.e.*;

$$p(\boldsymbol{x};\boldsymbol{\lambda}) \propto \exp\left[-\sum_{i=1}^{r}\lambda_i\phi_i(\boldsymbol{x})\right], \tag{12}$$

and find the conditions on the functions $\phi_i(\boldsymbol{x})$ for which $p(\boldsymbol{x})$ is scale invariant.

Note that these laws can be considered as the maximum entropy prior laws if our prior knowledge is:

- What we know about $\boldsymbol{x}$ is:

$$\mathrm{E}\left\{\phi_i(\boldsymbol{x})\right\} = d_i, \quad i = 1,\cdots,r,$$

- and what we know about the noise $\boldsymbol{b}$ is:

$$\left\{\begin{array}{l}\mathrm{E}\left\{\boldsymbol{b}\right\} = 0,\\ \mathrm{E}\left\{\boldsymbol{b}\boldsymbol{b}^t\right\} = \boldsymbol{R}_b = \sigma^2\boldsymbol{I},\end{array}\right.$$

  where $\boldsymbol{R}_b$ is the covariance matrix of $\boldsymbol{b}$.

Now, using the equations (11) and (12) and noting by $\boldsymbol{\theta} = (\sigma^2,\lambda_1,\cdots,\lambda_r)$, by $\boldsymbol{\lambda} = (\lambda_1,\cdots,\lambda_r)$, and by $\boldsymbol{\phi}(\boldsymbol{x}) = (\phi_1(\boldsymbol{x}),\cdots,\phi_r(\boldsymbol{x}))$, we have

$$p(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta}) \propto \exp\left[-\chi^2(\boldsymbol{x},\boldsymbol{y};\sigma^2) - \boldsymbol{\lambda}^t\boldsymbol{\phi}(\boldsymbol{x})\right], \tag{13}$$

and the scale invariance condition becomes:

$$\forall k>0, \forall \boldsymbol{x}_1,\boldsymbol{y}_1, \quad \chi_k^2(\boldsymbol{x}_k,\boldsymbol{y}_k;\sigma_k^2) + \boldsymbol{\lambda}_k^t\boldsymbol{\phi}(\boldsymbol{x}_k) = \chi_1^2(\boldsymbol{x}_1,\boldsymbol{y}_1;\sigma_1^2) + \boldsymbol{\lambda}_1^t\boldsymbol{\phi}(\boldsymbol{x}_1) + cte.$$

But with the Gaussian choice for the noise *pdf* we have

$$\forall k>0, \forall \boldsymbol{x}_1,\boldsymbol{y}_1,\ \chi_k^2(\boldsymbol{x}_k,\boldsymbol{y}_k;\sigma_k^2) = \frac{1}{2\sigma_k^2}\,||\boldsymbol{y}_k - \boldsymbol{H}\boldsymbol{x}_k||^2 = \frac{1}{2k^2\sigma_1^2}\,k^2\,||\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{x}_1||^2 = \chi_1^2(\boldsymbol{x}_1,\boldsymbol{y}_1;\sigma_1^2),$$

and so the condition becomes

$$\forall k>0, \forall \boldsymbol{x}, \quad \boldsymbol{\lambda}_k^t\boldsymbol{\phi}(\boldsymbol{x}_k) = \boldsymbol{\lambda}_1^t\boldsymbol{\phi}(\boldsymbol{x}_1) + cte, \tag{14}$$

or equivalently,

$$p_k(\boldsymbol{x}_k;\boldsymbol{\lambda}_k) = \frac{1}{k^n}\,p_1(\boldsymbol{x}_1;\boldsymbol{\lambda}_1) \quad \text{with} \quad \boldsymbol{\lambda}_k = f(\boldsymbol{\lambda}_1,k).$$

Thus, in the case of centered Gaussian *pdf* for the noise, to have a scale invariant posterior law it is sufficient to have a scale invariant prior law.

Now, assuming interchangeable (independent) pixels, *i.e.*;

$$p(\boldsymbol{x};\boldsymbol{\lambda}) = \exp\left[\lambda_0 + \sum_{i=1}^{r}\boldsymbol{\lambda}_i\phi_i(\boldsymbol{x})\right] = \prod_{j=1}^{N}p(x_j;\boldsymbol{\lambda}), \tag{15}$$

or equivalently,

$$\phi_i(\boldsymbol{x}) = \sum_{j=1}^{N} \phi_i(x_j) \tag{16}$$

we have to find the conditions on the scalar functions $\phi_i(x)$ of scalar variables x who satisfy the equation (14) or equivalently

$$\forall k > 0, \forall x, \quad \sum_{i=1}^{r} \lambda_i(k)\, \phi_i(kx) = \sum_{i=1}^{r} \lambda_i(1)\, \phi_i(x) + cte \tag{17}$$

We have shown (see appendix) that, the functions $\phi_i(x)$ which satisfy these conditions are all either the powers of x or the powers of $\ln x$ or a multiplication of them. The general expressions for these functions are:

$$\phi(x) = \sum_{m=1}^{M} \left(\sum_{n=0}^{N_m - 1} c_{mn} (\ln x)^n \right) x^{\alpha_m} + \sum_{n=0}^{N_0} c_{0n} (\ln x)^n, \quad \text{with } M \leq r \text{ and } \sum_{m=0}^{M} N_m = r \tag{18}$$

where M and N_m are integer numbers, and c_{mn}, c_{0n} and α_m are real numbers. For a geometrical interpretation and more details see appendix. The following examples show some special and interesting cases.

One parameter laws: Consider the case of $r = 1$. In this case we have

$$p(x; \lambda) \propto \exp\left[-\lambda \phi(x)\right]. \tag{19}$$

Applying the general rule with

$$r = 1 \longrightarrow \begin{cases} M = 0, N_0 = 1, & \longrightarrow c_{00} + c_{01} \ln x \\ M = 1, N_0 = 0, N_1 = 1, & \longrightarrow c_{00} + c_{10} x^{\alpha_1} \end{cases}$$

we find that the only functions who satisfy these conditions are:

$$\left\{ \phi(x) \right\} = \left\{ x^{\alpha}, \ln x \right\} \tag{20}$$

where α is a real number. There isc two interesting special cases:

- $\phi(x) = x^{\alpha}$, resulting to: $p(x) \propto \exp\left[-\lambda x^{\alpha}\right]$, $\alpha > 0, \lambda > 0$, which is a generalized Gaussian *pdf*, and
- $\phi(x) = \ln x$, resulting to: $p(x) \propto \exp\left[-\lambda \ln x\right]$, which is a special case of the Beta *pdf*.

Note that the famous *entropic* prior law: $p(x) \propto \exp\left[-\lambda x \ln x\right]$ of Gull and Skilling [11, 4] does not verify the scale invariance property. But, if we add one more parameter

$$p(x) \propto \exp\left[-\lambda x \ln x + \mu x\right],$$

then, it will satisfy this condition as we can see in the next section.

Two parameters laws: This is the case where $r = 2$ and we have:

$$p(x; \lambda) \propto \exp\left[-\lambda\phi_1(x) - \mu\phi_2(x)\right], \tag{21}$$

and applying the general rule:

$$r = 2 \longrightarrow \left\{ \begin{array}{ll} M = 2, N_0 = 0, N_1 = 1, N_2 = 1, & \longrightarrow c_{00} + c_{10}x^{\alpha_1} + c_{20}x^{\alpha_2} \\ M = 1, N_0 = 0, N_1 = 2, & \longrightarrow c_{00} + c_{10}x^{\alpha_1} + c_{11}x^{\alpha_1}\ln x \\ M = 1, N_0 = 1, N_1 = 1, & \longrightarrow c_{00} + c_{10}x^{\alpha_1} + c_{01}\ln x \\ M = 0, N_0 = 2, & \longrightarrow c_{00} + c_{01}\ln x + c_{02}\ln^2 x \end{array} \right.$$

we see that in this case the only functions (ϕ_1, ϕ_2) which satisfy these conditions are:

$$\Big\{(\phi_1(x), \phi_2(x))\Big\} = \quad \Big\{(x^{\alpha_1}, x^{\alpha_2}), (x^{\alpha_1}, x^{\alpha_1}\ln x), (x^{\alpha_1}, \ln x), (\ln x, \ln^2 x)\Big\} \tag{22}$$

where α_1 and α_2 are two real numbers. Special cases are obtained when we choose $\phi_2(x) = x$, the only possible functions for $\phi_1(x)$ are then:

$$\{x^\alpha, \ln x, x\ln x\}. \tag{23}$$

and we have the following interesting cases:

- $\phi_1(x) = x^2$, resulting to: $\quad p(x) \propto \exp\left[-\lambda x^2 - \mu x\right] \propto \exp\left[-\lambda\left(x + \frac{\mu}{2\lambda}\right)^2\right]$, which is a Gaussian *pdf* $\mathcal{N}\left(m = \frac{-\mu}{\lambda}, \sigma^2 = \frac{1}{2\lambda}\right)$.
- $\phi_1(x) = \ln x$, resulting to: $\quad p(x) \propto \exp\left[-\lambda\ln x - \mu x\right] = x^{-\lambda}\exp\left[-\mu x\right]$, which is the Gamma *pdf*, and finally,
- $\phi_1(x) = x\ln x$, resulting to: $\quad p(x) \propto \exp\left[-\lambda x\ln x - \mu x\right]$. which is known as the *entropic pdf*.

Three parameters laws: This is the case where $r = 3$. Once more applying the general rule we find:

$$r = 3 \rightarrow \left\{ \begin{array}{ll} M = 3, N_0 = 0, N_1 = 1, N_2 = 1, N_3 = 1, & \rightarrow c_{00} + c_{10}x^{\alpha_1} + c_{20}x^{\alpha_2} + c_{30}x^{\alpha_3} \\ M = 2, N_0 = 0, N_1 = 1, N_2 = 2, & \rightarrow c_{00} + c_{10}x^{\alpha_1} + c_{20}x^{\alpha_2} + c_{21}x^{\alpha_2}\ln x \\ M = 2, N_0 = 1, N_1 = 1, N_2 = 1, & \rightarrow c_{00} + c_{01}\ln x + c_{10}x^{\alpha_1} + c_{20}x^{\alpha_2} \\ M = 1, N_0 = 0, N_1 = 3, & \rightarrow c_{00} + c_{10}x^{\alpha_1} + c_{11}x^{\alpha_1}\ln x + c_{12}x^{\alpha_1}\ln^2 x \\ M = 1, N_0 = 1, N_1 = 2, & \rightarrow c_{00} + c_{01}\ln x + c_{10}x^{\alpha_1} + c_{11}x^{\alpha_1}\ln x \\ M = 1, N_0 = 2, N_1 = 1, & \rightarrow c_{00} + c_{01}\ln x + c_{02}\ln^2 x + c_{10}x^{\alpha_1} \\ M = 0, N_0 = 3, & \rightarrow c_{00} + c_{01}\ln x + c_{02}\ln^2 x + c_{03}\ln^3 x \end{array} \right.$$

which means:

$$\Big\{(\phi_1(x), \phi_2(x), \phi_3(x))\Big\} = \Big\{ \begin{array}{l} (x^{\alpha_1}, x^{\alpha_2}, x^{\alpha_3}),\ (x^{\alpha_1}, x^{\alpha_2}, \ln x),\ (x^{\alpha_1}, x^{\alpha_1}\ln x, x^{\alpha_1}\ln^2 x), \\ (x^{\alpha_1}, x^{\alpha_1}\ln x, \ln x),\ (x^{\alpha_1}, x^{\alpha_2}, x^{\alpha_2}\ln x),\ (x^{\alpha_1}, \ln x, \ln^2 x), \\ \hfill (\ln x, \ln^2 x, \ln^3 x)\Big\} \end{array} \tag{24}$$

where α_1, α_2 and α_3 are three real numbers.

4. Proposed method

The general procedure of the inversion method we propose can be resumed as follows:

- Choose a set of functions $\phi_i(x)$ between the possibles ones described in the last section and assign the prior $p(\boldsymbol{x})$. In many imaging applications we proposed and used successfully the following two parameters one:

$$p(\boldsymbol{x};\boldsymbol{\lambda}) \propto \exp\left[-\lambda_1 H(\boldsymbol{x}) - \lambda_2 S(\boldsymbol{x})\right], \quad \text{with } H(\boldsymbol{x}) = \sum_{j=1}^{N} \phi_1(x_j), \text{ and } S(\boldsymbol{x}) = \sum_{j=1}^{N} \phi_2(x_j)$$

  where $\phi_1(x)$ and $\phi_2(x)$ choosed between the possible ones in (22) or (23).
- When what we know about the noise $\boldsymbol{b}$ is only its covariance matrix $\mathrm{E}\left\{\boldsymbol{b}\boldsymbol{b}^t\right\} = \boldsymbol{R}_b = \sigma_b^2 \boldsymbol{I}$, then using the maximum entropy principle we have:

$$p(\boldsymbol{y}|\boldsymbol{x}) \propto \exp\left[-\frac{1}{2}Q(\boldsymbol{x})\right], \quad \text{with } Q(\boldsymbol{x}) = (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^t {\boldsymbol{R}_b}^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}).$$

  We may note that $p(\boldsymbol{y}|\boldsymbol{x})$ is also a scale invariant probability law.
- Using the Bayes' rule and MAP estimator the solution is determined by

$$\widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}}\left\{p(\boldsymbol{x}|\boldsymbol{y})\right\} = \arg\min_{\boldsymbol{x}}\left\{J(\boldsymbol{x})\right\}, \quad \text{with } J(\boldsymbol{x}) = Q(\boldsymbol{x}) + \lambda_1 H(\boldsymbol{x}) + \lambda_2 S(\boldsymbol{x}).$$

  Note here also that, for the cases where one of the functions $\phi_1(x)$ or $\phi_2(x)$ is a logarithmic function of x, we have to constraint its range to the positive real axis, and we have to solve the following optimization problem

$$\widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}>0}\left\{p(\boldsymbol{x}|\boldsymbol{y})\right\} = \arg\min_{\boldsymbol{x}>0}\left\{J(\boldsymbol{x})\right\}.$$

  This optimization is achieved by a modified conjugate gradients method.
- The choice of the functions $\phi_i(x)$ and the determination of the parameters (λ_1, λ_2) in the first step is still an open problem.
  In imaging applications we propose to do this choice from our prior knowledge on the nature of interested quantity (physics of the application). For example, if the object $\boldsymbol{x}$ is a real quantity equally distributed on the positive and the negative reals then a Gaussian prior, *i.e.*; $(\phi_1(x) = x, \phi_2(x) = x^2)$ is convenient. But, if the object $\boldsymbol{x}$ is a positive quantity or if we know that it represents small extent, bright and sharp objects on a nearly black background (images in radio astronomy, for example), then we may choose $(\phi_1(x) = x, \phi_2(x) = \ln x)$, or $(\phi_1(x) = x, \phi_2(x) = x\ln x)$ which are the priors with longer tails than the Gaussian or truncated Gaussian one.
  When the choice of the functions $(\phi_1(x), \phi_2(x))$ is done, we still have to determine the hyperparameters (λ_1, λ_2). For this two main approaches have been proposed. The first is based on the generalized maximum likelihood (GML) which tries to estimate simultaneously the parameters $\boldsymbol{x}$ and the hyperparameters $\boldsymbol{\theta} = (\lambda_1, \lambda_2)$ by

$$(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{x},\boldsymbol{\theta})}\left\{p(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})\right\} = \arg\max_{(\boldsymbol{x},\boldsymbol{\theta})}\left\{p(\boldsymbol{y}|\boldsymbol{x})\,p(\boldsymbol{x};\boldsymbol{\theta})\right\}, \qquad (25)$$

and the second is based on the marginalization (MML), in which the hyperparameters $\boldsymbol{\theta}$ are estimated first by

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ p(\boldsymbol{y};\boldsymbol{\theta}) = \int p(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{x} \right\} = \arg\max_{\boldsymbol{\theta}} \left\{ \int p(\boldsymbol{y}|\boldsymbol{x})\,p(\boldsymbol{x};\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{x} \right\}, \tag{26}$$

and then used for the estimation of $\boldsymbol{x}$:

$$\widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \left\{ p(\boldsymbol{x}|\boldsymbol{y};\widehat{\boldsymbol{\theta}}) \right\} = \arg\max_{\boldsymbol{x}} \left\{ p(\boldsymbol{y}|\boldsymbol{x})\,p(\boldsymbol{x}|\widehat{\boldsymbol{\theta}}) \right\}. \tag{27}$$

What is important here is that both methods preserve the scale invariant property. For practical applications we have recently proposed and used a method based on the generalized maximum likelihood [8, 9] which has been successfully used in many signal and image reconstruction and restoration problems as we mentionned in the introduction [10].

5. Conclusions

Excepted the Gaussian case where all the Bayesian estimators are linear functions of the observed data, in general, the Bayesian estimators are *nonlinear* functions of the data. When dealing with linear inverse problems linearity is sometimes a too strong property, while *scale invariance* often remains a desirable property. In this paper we discussed and proposed a familly of generalized exponential probability distributions for the direct probabilities (the prior $p(\boldsymbol{x})$ and the likelihood $p(\boldsymbol{y}|\boldsymbol{x})$), for which the posterior $p(\boldsymbol{x}|\boldsymbol{y})$, and, consequently, the main posterior estimators are scale invariant. Among many properties, generalized exponential can be considered as the maximum entropy probability distributions subject to the knowledge of a finite set of expectation values of some knwon functions.

1. Appendix: General case

We want to find the solutions of the following equation:

$$\forall k > 0, \forall x, \quad \sum_{i=1}^{r} \lambda_i(k)\phi_i(kx) = \sum_{i=1}^{r} \lambda_i(1)\phi_i(x) + \beta(k) \tag{A.1}$$

Making the following changes of variables and notations

$$1/k = \tilde{k},\ kx = \tilde{x},\ \lambda_i(k) = \tilde{\lambda}_i(\tilde{k}), \quad \text{and} \quad \beta_i(k) = \tilde{\beta}_i(\tilde{k}),$$

equation (A.1) becomes

$$\sum_{i=1}^{r} \tilde{\lambda}_i(\tilde{k})\phi_i(\tilde{x}) = \sum_{i=1}^{r} \tilde{\lambda}_i(1)\phi_i(\tilde{k}\tilde{x}) + \tilde{\beta}(\tilde{k})$$

For convenience sake, we will drop out the tilde ˜, and note $\lambda_i(1) = \lambda_i$, so that we can write

$$\sum_{i=1}^{r} \lambda_i(k)\phi_i(x) = \sum_{i=1}^{r} \lambda_i\phi_i(kx) + \beta(k)$$

Noting

$$S(x) = \sum_{i=1}^{r} \lambda_i \phi_i(x), \quad \text{and so} \quad S(kx) = \sum_{i=1}^{r} \lambda_i \phi_i(kx)$$

we have

$$\sum_{i=1}^{r} \lambda_i(k)\phi_i(x) = S(kx) + \beta(k) \tag{A.2}$$

Deriving $r-1$ times this equation with respect to k we obtain

$$\begin{aligned}
\sum_{i=1}^{r} \lambda_i'(k)\phi_i(x) &= x\,S'(kx) + \beta'(k) \\
\sum_{i=1}^{r} \lambda_i''(k)\phi_i(x) &= x^2\,S''(kx) + \beta''(k) \\
\vdots \qquad\qquad &\quad \vdots \\
\sum_{i=1}^{r} \lambda_i^{(r-1)}(k)\phi_i(x) &= x^{r-1} S^{(r-1)}(kx) + \beta^{(r-1)}(k)
\end{aligned} \tag{A.3}$$

Combinig equations (A.2) and (A.3) in matrix form we have

$$\begin{pmatrix} \lambda_1(k) & \cdots & \lambda_r(k) \\ \lambda_1'(k) & \cdots & \lambda_r'(k) \\ \lambda_1''(k) & \cdots & \lambda_r''(k) \\ \vdots & \cdots & \vdots \\ \lambda_1^{(r-1)}(k) & \cdots & \lambda_r^{(r-1)}(k) \end{pmatrix} \begin{pmatrix} \phi_1(x) \\ \phi_2(x) \\ \phi_3(x) \\ \vdots \\ \phi_r(x) \end{pmatrix} = \begin{pmatrix} S(kx) + \beta(k) \\ xS'(kx) + \beta'(k) \\ x^2 S''(kx) + \beta''(k) \\ \vdots \\ x^{r-1} S^{(r-1)}(kx) + \beta^{(r-1)}(k) \end{pmatrix} \tag{A.4}$$

If this matrix equation can be inverted, this means that any function $\phi_i(x)$ is a linear combination of $S(kx) + \beta(k)$ and its $(r-1)$ derivatives with respect to k:

$$\phi_i(x) = \sum_{i=0}^{r} \eta_i(k) \left[x^{(i-1)} S^{(i-1)}(kx) + \beta^{(i-1)}(k) \right], \tag{A.5}$$

and if this is not the case, this means that there exists an interval for k, for which some of the functions $\lambda_i(k)$ are linear combinations of the others [2]. In this case let us show that we will go back to the situation of the problem of lower order r. Let us to assume that the last column of the matrix is a linear combination of the others, *i.e.*;

$$\lambda_r(k) = \sum_{i=1}^{r-1} \gamma_i \lambda_i(k).$$

Putting this in the equation (A.1) will give

$$\sum_{i=1}^{r-1} \lambda_i(k)\phi_i(kx) + \left[\sum_{i=1}^{r-1} \gamma_i \lambda_i(k) \right] \phi_r(kx) = \sum_{i=1}^{r-1} \lambda_i(1)\phi_i(x) + \beta(k) + \left[\sum_{i=1}^{r-1} \gamma_i \lambda_i(1) \right] \phi_r(x)$$

and noting $\psi_i(x) = \phi_i(x) + \gamma_i \phi_r(x)$ and $\psi_i(kx) = \phi_i(kx) + \gamma_i \phi_r(kx)$ we obtain

$$\sum_{i=1}^{r-1} \lambda_i(k) \psi_i(kx) = \sum_{i=1}^{r} \lambda_i(1) \psi_i(x) + \beta(k)$$

which is an equation in the same form of (A.1), but of lower order.

Deriving now both parts of the equation (A.5) with respect to k and noting $kx = u$ we obtain

$$\sum_{i=0}^{r} a_i \, u^i S^i(u) = a \tag{A.6}$$

This is the general expression of a rth order Euler–Cauchy differential equation [1, 2] which is classically solved through the change of variable $u = e^x$, and one can find the general expression of its solution in the following form:

$$S(x) = \sum_{m=1}^{M} \left(\sum_{n=0}^{N_m - 1} c_{mn} (\ln x)^n \right) x^{\alpha_m} + \sum_{n=0}^{N_0} c_{0n} (\ln x)^n \quad \text{with } M = 0, \cdots r, \quad \text{and} \quad \sum_{m=0}^{M} N_m = r \tag{A.7}$$

where M and N_m are integer numbers, and c_{mn}, c_{0n} and α_m are real numbers. In fact the most general solution also incorporate terms of the form

$$\left[\sum_n (\ln x)^n \left(\alpha_n \cos(\ln x) + \beta_n \sin(\ln x) \right) \right] x^d$$

derived from complex α_m and c_{mn}. But we will not consider these terms because the resulting *pdf*'s have oscillatory behavior around zero.

One can give a geometric interpretation of the solutions given in (A.7). For any given order r make a $(r+1) \times (r+1)$ table in the form

$\ln^r x$					
$\vdots$					
$\ln^2 x$					
$\ln x$					
1	$\times$				
	1	x^{α_1}	x^{α_2}	$\cdots$	x^{α_r}

and let r mass points fall down into the columns. To each filled box is assigned a function $\phi_i(x)$ by multiplying the corresponding powers of x and $\ln x$ on the same line and the same column. To illustrate this, we give in the following the three first cases:

<table>
<tr><th>Case $r = 1$:</th><th>Case $r = 2$:</th><th>Case $r = 3$:</th></tr>
<tr>
<td>$\begin{array}{c|c|c} \ln x & b & \\ \hline 1 & \times & a \\ \hline & 1 & x^{\alpha_1} \end{array}$</td>
<td>$\begin{array}{c|c|c|c} \ln^2 x & d & & \\ \hline \ln x & bd & c & \\ \hline 1 & \times & abc & a \\ \hline & 1 & x^{\alpha_1} & x^{\alpha_2} \end{array}$</td>
<td>$\begin{array}{c|c|c|c|c} \ln^3 x & g & & & \\ \hline \ln^2 x & fg & c & & \\ \hline \ln x & bdfg & dc & e & \\ \hline 1 & \times & abcdef & abe & a \\ \hline & 1 & x^{\alpha_1} & x^{\alpha_2} & x^{\alpha_3} \end{array}$</td>
</tr>
<tr>
<td>$\begin{array}{c|c} & \phi(x) \\ \hline a & x^{\alpha_1} \\ b & \ln x \end{array}$</td>
<td>$\begin{array}{c|cc} & \phi_1(x) & \phi_2(x) \\ \hline a & x^{\alpha_1} & x^{\alpha_2} \\ b & x^{\alpha_1} & \ln x \\ c & x^{\alpha_1} & x^{\alpha_1} \ln x \\ d & \ln x & \ln^2 x \end{array}$</td>
<td>$\begin{array}{c|ccc} & \phi_1(x) & \phi_2(x) & \phi_3(x) \\ \hline a & x^{\alpha_1} & x^{\alpha_2} & x^{\alpha_3} \\ b & x^{\alpha_1} & x^{\alpha_2} & \ln x \\ c & x^{\alpha_1} & x^{\alpha_1} \ln x & x^{\alpha_1} \ln^2 x \\ d & x^{\alpha_1} & x^{\alpha_1} \ln x & \ln x \\ e & x^{\alpha_1} & x^{\alpha_2} & x^{\alpha_2} \ln x \\ f & x^{\alpha_1} & \ln x & \ln^2 x \\ g & \ln x & \ln^2 x & \ln^3 x \end{array}$</td>
</tr>
</table>

References

1. Angot A., “Compléments de mathématiques,” Masson ed., Sixième Édition, Paris, 1982.
2. Bass J., “Cours de mathématiques,” Masson ed., Tome II, Quatrième Édition, Paris, 1968.
3. Demoment G., “Image Reconstruction and Restoration: Overview of Common Estimation Structure and Problems,” *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 37, pp:2024-2036, (1989).
4. Gull S. F. and Skilling J., “Maximum entropy method in image processing,” *IEE Proc.*, **131-F**, pp. 646-659, 1984.
5. Mohammad-Djafari A. and Idier J., “Maximum entropy prior laws of images and estimation of their parameters,” in *T.W. Grandy (ed.), Maximum-entropy and Bayesian methods,* Kluwer Academic Publishers, Netherlands, 1990.
6. Mohammad-Djafari A. and Demoment G., “Maximum entropy Fourier synthesis with application to diffraction tomography,“ *Applied Optics*, Vol.26, No. 10, pp:1745-1754, (1987).
7. Mohammad-Djafari A. and Demoment G., “Maximum entropy reconstruction in X ray and diffraction tomography,” *IEEE Trans. on Medical Imaging*, Vol. 7, No. 4 pp:345-354, (1988).
8. Mohammad-Djafari A., “Bayesian Approach with Maximum Entropy Priors to Imaging Inverse Problems, Part I: Fundations,” *submitted to IEEE Trans. on Image Processing*, (August, 1993).
9. Mohammad-Djafari A., “Bayesian Approach with Maximum Entropy Priors to Imaging Inverse Problems, Part II: Applications,” *submitted to IEEE Trans. on Image Processing*, (August, 1993).
10. Nguyen M.K. and Mohammad-Djafari A., “Bayesian Maximum Entropy Image Reconstruction from the Microwave Scattered Field Data,” in *A. Mohammad-Djafari and G. Demoment(ed.), Maximum Entropy and Bayesian Methods,* Kluwer Academic Publishers, the Netherlands, 1993.
11. Skilling J., “Maximum-Entropy and Bayesian Methods,” J. Skilling ed., Dordrecht: Kluwer Academic Publisher, 1988.